Carmeliet, P. (2003). Nat. Med. *9*, 653–660.

Carmeliet, P., and Luttun, A. (2001). Thromb. Haemost. *86*, 289–297.

Carmeliet, P., and Tessier-Lavigne, M. (2005). Nature *436*, 193–200.

Ceradini, D.J., Kulkarni, A.R., Callaghan, M.J., Tepper, O.M., Bastidas, N., Kleinman, M.E., Capla, J.M., Galiano, R.D., Levine, J.P., and Gurtner, G.C. (2004). Nat. Med. *10*, 858–864.

De Palma, M., Venneri, M.A., Galli, R., Sergi, L.S., Politi, L.S., Sampaolesi, M., and Naldini, L. (2005). Cancer Cell *8*, 211–226.

Foo, S.E., Turner, C.J., Adams, S., Compagni, A., Aubyn, D., Kogata, N., Lindblom, P., Shani, M., Zicha, D., and Adams, R.H. (2006). Cell, this issue.

Grunewald, M., Avraham, I., Dor, Y., Bachar-Lustig, E., Itin, A., Yung, S., Chimenti, S., Landsman, L., Abramovitch, R., and Keshet, E. (2006). Cell, this issue.

Orimo, A., Gupta, P.B., Sgroi, D.C., Arenzana-Seisdedos, F., Delaunay, T., Naeem, R., Carey, V.J., Richardson, A.L., and Weinberg, R.A. (2005). Cell *121*, 335–348.

Quesenberry, P.J., Colvin, G., and Abedi, M. (2005). Exp. Hematol. *33*, 9–19.

Rafii, S., and Lyden, D. (2003). Nat. Med. *9*, 702–712.

Takakura, N., Watanabe, T., Suenobu, S., Yamada, Y., Noda, T., Ito, Y., Satake, M., and Suda, T. (2000). Cell *102*, 199–209.

Urbich, C., and Dimmeler, S. (2004). Circ. Res. *95*, 343–353.

Transcription Factor Target Practice

Frank C.P. Holstege[1,*] and Hans Clevers[2]
[1]Department of Physiological Chemistry, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, the Netherlands
[2]Hubrecht Laboratory and Centre for Biomedical Genetics, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands
*Contact: f.c.p.holstege@med.uu.nl
DOI 10.1016/j.cell.2005.12.026

In this issue of *Cell*, Hallikas et al. (2006) and Wei et al. (2006) describe different ways of identifying direct targets of transcription factors and their corresponding regulatory sequences in the genome. Although still under development, these studies provide an efficient way to decipher regulatory networks.

The identification of downstream targets of regulatory factors is required to understand cellular responses to environmental and developmental cues. Precise control of gene expression is achieved through regulators such as microRNAs and transcription factors (He and Hannon, 2004; Pabo and Sauer, 1992). Which genes are controlled by gene-specific transcription factors is partly determined by the DNA binding domain of these proteins. This domain allows the transcription factor to bind to specific DNA motifs in the vicinity of target genes. In theory, the availability of whole genome sequences should have made the job of finding transcription factor targets straightforward, for example, by allowing the location of recognition motifs to be simply looked up in the genome sequence. In practice, the task of determining transcription factor targets is still daunting, especially for organisms with complex genomes. Two papers in this issue of *Cell* (Hallikas et al., 2006; Wei et al., 2006) describe different ways of locating direct targets of transcription factors and their corresponding regulatory regions. The studies focus on transcription factors that are important for development and oncogenesis. The results underscore the importance of such work and indicate that there is still much to be discovered about even well-characterized transcription factors. Additionally, these results emphasize that combinations of different approaches will be required for the complete determination of regulatory networks.

Prior to the availability of whole genome sequences, methods for determining transcription factor targets were cumbersome. Such studies usually began by studying easily accessible proximal promoter sequences of a candidate target gene. Important regulatory DNA motifs were determined by mutating promoter sequences (promoter-bashing). This was followed by arduous hit-or-miss approaches such as electrophoretic mobility shift assays, biochemical purification, or expression cloning to identify the corresponding transcription factors. The availability of whole genome sequences has reversed this process and has made it possible to develop more exhaustive search methods.

The first contribution of whole genome sequences for finding transcription factor targets has been the development of microarray technology to measure mRNA expression (Young, 2000). This can identify all genes that exhibit significant changes in mRNA levels upon inactivation of a transcription factor. The degree to which all direct targets are found depends on functional redundancy and whether the particular developmental stage or growth condition being studied completely covers the role of the transcription factor. The portion of such genes that actually represents direct targets is dependent on the number and the nature of indi-

rect effects. For the purpose of finding direct transcription factor targets, microarray expression profiling on its own is insufficient. However, because the readout of microarray expression-profiling experiments (mRNA levels) is close to the process of interest (transcription), it is the best way of characterizing the amazing spectrum of regulatory changes that occur during diverse cellular programs.

A more direct way in which the availability of whole genome sequences can in principle contribute to finding transcription factor targets is by searching for the corresponding recognition motif within the genome sequence itself. This approach is taken by Hallikas et al. (2006). Previous endeavors have made it clear that two major hurdles need to be overcome, especially for complex genomes. First, the recognition motif needs to be well defined as knowledge of a single binding site is not sufficient. DNA sequence motifs generally have a "loose" consensus that can only be determined by comparison of a large collection of binding sites. One significant aspect of the current study is the initial determination of the DNA binding specificities of the transcription factors in order to yield a consensus binding site for searching the genome.

The second major hurdle is the search space itself. Mammalian genomes are enormous compared to the six to eight base-pair DNA motifs generally recognized by transcription factors. Also, because consensus binding motifs are loose, the number of putative binding sites vastly exceeds the number of functional sites. One way of reducing the search space is to restrict the analysis to promoters, relatively short regions of regulatory DNA immediately upstream of gene coding regions. This has several disadvantages. It relies on accurate genome annotation, which has not been achieved yet for all genes or noncoding RNAs in most genomes. Furthermore, it excludes finding binding sites in enhancer regions that can be located many tens of kilobases away from the transcribed DNA under its control (Hertel et al., 1997). Hallikas et al. (2006) have circumvented such disadvantages by making use of two additional properties of regulatory DNA sequences: its conserved nature and the fact that on such regulatory regions the binding sites of many different transcription factors tend to cluster together. The authors compare orthologous regulatory sequences based on aligning transcription factor binding sites rather than the sequence itself and have also introduced a penalty score for sequences inserted between binding sites that may compromise cooperative binding. All of these features have been combined into a scoring scheme algorithm called EEL (*e*nhancer *e*lement *l*ocator), which is capable of identifying conserved enhancer elements. Another important feature of this study is the experimental validation of enhancer activity for several of their in silico predictions.

A more direct approach for identification of targets is to determine transcription factor location on DNA in vivo by chromatin immunoprecipitation (ChIP) (Orlando et al., 1997). This technique uses antibodies against specific transcription factors to isolate the bound DNA sequences embedded in chromatin. ChIP can be coupled to microarray technology (Ren et al., 2000). Although representation of entire genomes is feasible, a large number of arrays are currently required to completely cover the largest genomes properly. The approach taken by Wei et al. (2006) circumvents this problem by coupling ChIP with paired-end ditag sequencing (ChIP-PET). This method extracts two 18 bp sequence tags, one from each end of a DNA fragment isolated by ChIP, and joins these two sequences together. These concatenated tags are then sequenced to identify transcription factor binding sites. The relatively unbiased nature of this approach makes it more favorable than single array studies with incomplete genome coverage. In the long run it is likely that the ChIP-PET technique will be replaced by arrays with a higher number of features than is currently feasible.

So what do these new technologies deliver? The EEL tool (Hallikas et al., 2006) correctly predicts several well-known enhancers in *Drosophila* and mammals. In a selected set of 13 known Wnt/Tcf4 targets, a significant enrichment of Tcf4 cognate sequences was seen in the predicted enhancer elements near these genes. A genome-wide search for Tcf4-driven enhancer elements identified 132 candidates. A total of six mapped close to previously reported Tcf target genes, which seems a low number given the hundreds of published Tcf target genes (for example, http://www.stanford.edu/~rnusse/pathways/targets.html). Expression analysis revealed that out of 12 genes analyzed, 5 were expressed in at least one embryonic site where Wnt regulation is known to occur. Two Tcf4-driven enhancers were predicted near the Wnt target gene *c-myc*. Both enhancers were shown to control tissue-specific transcription of reporter genes in transgenic mice, but regulation by Wnt/Tcf4 was not investigated.

A genome-wide search for enhancer elements containing binding sites for the Hedgehog transcription factor GLI identified 42 candidates including two out of the three in vivo validated Hedgehog targets. For 16 other predicted GLI-controlled enhancers, the expression of the associated genes was determined. For five of these genes, the expression patterns were consistent with regulation by Hedgehog. Three out of four predicted that GLI-driven enhancers exerted control over the expression of reporter genes in transgenic mice. Additionally, two out of four potential GLI-controlled enhancers in the *N-myc* gene acted as transgenic enhancers in vivo, although whether Hedgehog/GLI signaling was involved was not addressed. Taken together, EEL appears to be a powerful tool to identify enhancer regions. However, its strength to make more specific predictions, such as control by Hedgehog or Wnt, awaits further experimental confirmation.

Ruan and colleagues applied their ChIP-PET method to identify p53 targets in a colorectal cancer cell line in which p53 is activated. In total, 1776 potential p53 targets were found more than once in the over 0.5 million sequenced ditags. When these p53 target genes were compared with a

list of 66 previously proposed p53 target genes, 41 were found on both lists. Further ChIP analysis indicated that only three of the "missed" genes in this set of 66 actually contain p53 binding sites, implying that the ditag approach produced few false negatives. Moreover, when 40 sequences were randomly selected from a higher confidence ditag list, all could be confirmed by conventional ChIP. The study had predictive value; from the collection of targets, the consensus p53 binding site could be refined and almost one hundred new p53 targets are proposed. Previously, a ChIP array-based study had assigned p53 binding sites to chromosomes 21 and 22, using the same cell line, albeit under different experimental conditions (Cawley et al., 2004). It is somewhat disconcerting to note that there is little overlap between the sets of proposed p53 binding sites from these two studies.

Besides the importance for understanding individual transcription factors, such genome-wide approaches are pivotal for deciphering regulatory networks, one of the principal goals of what is currently called systems biology. The two types of approaches are complementary. ChIP-based approaches are more direct but are limited by the particular developmental stage or growth condition analyzed. Bioinformatic predictions are perhaps better able to identify all possible regulatory regions, but these then need more rigorous testing for activity and transcription factor binding. The EEL approach relies in part on orthologous regulatory regions and assumes that binding site specificities will also be conserved. The degree of both assumptions has not yet been clearly established. Only a subset of transcription factor binding sites are presently included in the EEL analysis. Expansion will therefore require determination of more consensus binding sites. This may be better driven by the results of ChIP studies than by in vitro selection of binding sites. Both methods, as well as array-based ChIP, require further development. However, in comparison to old-fashioned "promoter-bashing," these studies provide efficient tools for deciphering regulatory networks in the postgenomic era.

REFERENCES

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. (2004). Cell *116*, 499–509.

Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E., and Taipale, J. (2006). Cell, this issue.

He, L., and Hannon, G.J. (2004). Nat. Rev. Genet. *5*, 522–531.

Hertel, K.J., Lynch, K.W., and Maniatis, T. (1997). Curr. Opin. Cell Biol. *9*, 350–357.

Orlando, V., Strutt, H., and Paro, R. (1997). Methods *11*, 205–214.

Pabo, C.O., and Sauer, R.T. (1992). Annu. Rev. Biochem. *61*, 1053–1095.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Science *290*, 2306–2309.

Young, R.A. (2000). Cell *102*, 9–15.

Wei, C.L., Wu, Q., Vega, V., Chu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y.T., Weng, Z., et al. (2006). Cell, this issue.

RISC-y Memories

Erica White-Grindley[1] and Kausik Si[1,*]

[1]The Stowers Institute for Medical Research, 1000 E 50th St., Kansas City, MO 64110, USA

*Contact: ksi@stowers-institute.org

DOI 10.1016/j.cell.2005.12.027

Local protein synthesis in the synapse is required for synaptic plasticity and has been implicated in learning and memory. However, direct evidence that behavioral training induces local protein synthesis has been lacking. In this issue of *Cell*, Ashraf et al. (2006) observe persistent local protein synthesis in the antennal lobe synapses of the fruit fly following olfactory-avoidance learning. This protein synthesis is regulated by the RNA-induced silencing complex (RISC).

Sensory experiences alter the electrical properties of ensembles of neurons. Retention of this altered state of neuronal activity is thought to constitute memory. Work on both vertebrates and invertebrates suggests that this alteration of neuronal properties is partly due to the change in the molecular composition of the synapse activated by a particular pattern of activity. Proteins can be made locally in the synapse in addition to the cell body, and this local synthesis can lead to synapse-specific changes in molecular composition (Steward and Schuman, 2001).

Previous studies suggested that local protein synthesis in the synapse has at least two distinct functions: synthesis of retrograde messengers that travel from the synapse to the cell body to activate mRNA synthesis and "marking" of the activated synapse to selectively use the proteins